Intelligent Information Fusion in the Aviation Domain: A Semantic-Web based Approach

Naveen Ashish [*] and Andre Goforth
{ashish@email.arc.nasa.gov, agoforth@mail.arc.nasa.gov}
NASA Ames Research Center, Moffett Field, CA, 94035-1000, USA

Information fusion from multiple sources is a critical requirement for envisioned "System Wide Information Management" in the National Airspace (NAS). NASA and the FAA envision creating an "integrated pool" of information originally coming from different sources, into which users, intelligent agents and NAS decision support tools (DSS) can tap into. In this paper we present the results of our initial investigations into the requirements and prototype development of such an integrated information pool for the NAS. We have attempted to ascertain key requirements for such an integrated pool based on a survey of DSS tools that will benefit from this integrated information pool. We then advocate key technologies from computer science research areas such as the semantic web, information integration, and intelligent agents that we believe are well suited to achieving the envisioned system wide information management capabilities.

I.   Introduction

Consider a situation where bad weather around a particular airport requires that the controllers suspend landing and take-off of any aircraft from that airport for a 2 hour period. All flights scheduled to land into or take-off from this airport in that specified period then need to be rerouted or rescheduled. Naturally this disruption causes flight delays, inconveniences air passengers, and may result in additional expenses for the airlines and the FAA. Can we deal *efficiently* and *optimally* with such situations i.e., can we do the required re-scheduling and rerouting by minimizing the flight delays and missed connections and also minimizing the additional financial costs to the airlines and authorities ? This is precisely the goal of the NASA FAA System Wide Information Management (SWIM) initiative. The SWIM initiative proposes to have various distributed information sources and services in National Airspace (NAS) connected seamlessly so that end users such as air-traffic controllers, other decision makers and decision aids such as decision support tools (DSS) or intelligent agents have integrated access to all the information needed to make informed decisions in the above kinds of situations. In this paper we present the results of our initial investigation into a critical aspect of SWIM which is the information integration issue. Seamless integrated access needs to be provided to the various distributed, heterogeneous and autonomously created and maintained information sources and services in the NAS. One of our thrusts has been to ascertain the information integration requirements in the NAS domain by talking to various potential stake-holders and familiarizing ourselves with the information access needs in the NAS. The other thrust has been to then evaluate what concepts, technologies, approaches and systems we can take from the state-of-the-art in computer science and information management research to apply to the information integration problem in the NAS. This has been with the goal of developing a prototype Aviation Information Integration and Monitoring System (AIMS) demonstrating the information access and integration capabilities. We present our findings on the above in the following sections concluding with recommendations for further activity.

II.   The NAS System and DSS Tools

We began our effort by first obtaining a preliminary understanding of the National Airspace System NAS and Air Traffic Management (ATM). We strongly recommend the reader to look up articles such as (Green et al. 2000) which provide a comprehensive and well written introduction to the NAS and ATM. It is clearly stressed that it is the NAS *DSS* tools that would benefit greatly by having access to an integrated pool of information (as opposed to

---

[*] With the USRA Research Institute for Advanced Computer Science at NASA Ames Research Center

accessing the information through multiple stovepipe connections to the various information sources and services) they need to access. We thus set for ourselves the following goals to better understand the domain.

- Obtain an understanding of NAS DSS tools, especially from the perspective of their requiring integrated access to multiple data sources.
- Determine the various types of data sources that are used by DSS tools and that need to be integrated into the envisioned SWIM information pool.

We interviewed a number of personnel in the NASA Ames Advanced Air Transportation Technologies (AATT) office who are intimately familiar with such NAS DSS tools and have actually developed some of these tools in the past. We conducted personal interviews and also surveyed a number of Operational Concept Documents (OCDs) for these DSS tools as advised by them. Some of the key information that we have ascertained is as follows:

- Information in the NAS comes from a wide *variety* of information sources and is of different kinds.
  The information could be time referenced information, geographically referenced information (such as airport locations, terrain information etc), weather information, air-traffic hazard information (bird activity), navigation system information, flight information, safety information (accident statistics etc) and other business management information.

- There are different kinds of *systems* providing and accessing information
  Namely flight service station systems, tower systems, terminal systems, enroute systems, oceanic systems, TFM (Traffic Flow Management) systems, and security systems.

- DSS tools fall into various categories based on the high level problems they were designed to address
  The high level goals for the future NAS include providing emerging situational awareness, collaborative decision making, and moving towards "free-flight" concepts. The key problems in the NAS that they are targeted to address are:

    1. Airspace organization and management.
    2. Demand capacity and balancing.
    3. Traffic synchronization, and
    4. Separation management .

  DSS tools fall into the following 7 broad categories:

  (a) Surface DSS
  (b) Terminal DSS
  (c) Enroute DSS
  (d) Oceanic DSS
  (e) TFM DSS
  (f) Infrastructure Management DSS
  (g) Flight Planning DSS

With guidance from NASA Ames AATT office personnel, the AIMS team surveyed a number of Operational Concept Documents (OCDs) for NAS DSS Tools. Some of the relevant tool descriptions we surveyed were the Surface Management System (SMS), Traffic Flow Management system (TFM) and the Surface Movement Advisor (SMA). Table 1 provides a summary outline description and the information sources accessed by these tools.

DSS Tool	Description and capabilities	Information Sources/Information Accessed

Surface Management System (SMS)	NASA FAA collaboration to develop a DSS tool. SMS (NASA 2003) helps controllers and air carriers collaboratively manage the movements of aircraft on the surface of busy airports, thereby improving capacity, efficiency, and flexibility.	(a) ETMS [†]
		   - Planned pushback times, flight plans, EDCTs (Expected Departure Clearance Time)
		(b) Surface Surveillance
		
		(c) Airport configuration, Miles-in-trail (National Log Program)
		(d) Ramp Tower
		   - Parking gates, pushback events, ready to push status, flight priorities

Traffic Flow Management (TFM)	Air traffic management is the strategic management of traffic flow to minimize delays and congestion while maximizing the overall throughput of the NAS. The goal of TFM is to provide strategic planning and management of air traffic demand to ensure smooth and efficient traffic flow through FAA-controlled airspace.	(a) ETMS
		   - GS/GDP parameters, schedules and flight plans, SUAs, radar tracks, etc.
		(b) ATCSCC (Air Traffic Control System Command Center)
		   - Routes, advisories, equipment status
		   - Demand rate, alert thresholds, reroutes, capacity
		(c) ARTCCs TMU (Air Route Traffic Control Center, Traffic Management Unit)
		   - NAS messages, flow rates, thresholds etc.
		(d) AOCs (Airline Operations Control Center)
		   - Flight data messages, cancellations, substitutions, candidate flights

Surface Movement Advisor (SMA)	The SMA, a joint development between the FAA and NASA, is designed to assist in airport surface operations. SMA electronically connects ATC, airline, and airport operations information to facilitate information sharing and improve taxi queuing. Further integration is also required with ground tracking systems.	(a) Controller inputs from the ramp and FAA towers (b) Live terminal area data from the FAA's HOST computer system (c) Airline schedules (d) Estimated pushback times, Gate assignments

Table 1. Representative DSS Tools

The reader is referred to http://www.as.nasa.gov/aatt/research.html#72 where a number of OCDs describing the above mentioned and several other tools are available.

III.   Technologies

The information sharing and integration problem is ubiquitous and indeed the computer science and information management research community as well as the software industry have contributed approaches, concepts, techniques, tools and products towards addressing the problem or aspects of it. Some of the technical areas that are highly relevant include information integration systems, intelligent agents, semantic web technologies such as mark-up languages and ontologies, information retrieval, etc. While we cannot go into detailed introductory descriptions

---

[†] Enhanced Traffic Management System (http://hf.tc.faa.gov/projects/etms.htm)

of these technologies we would like to point the reader to (Knoblock et al. 2001) for an excellent introduction to intelligent information agents and to (Berners-Lee et al. 2001) for an introductory tutorial of the concept of the semantic web. We now present our evaluation of technologies in these areas for the SWIM information extraction and access needs.
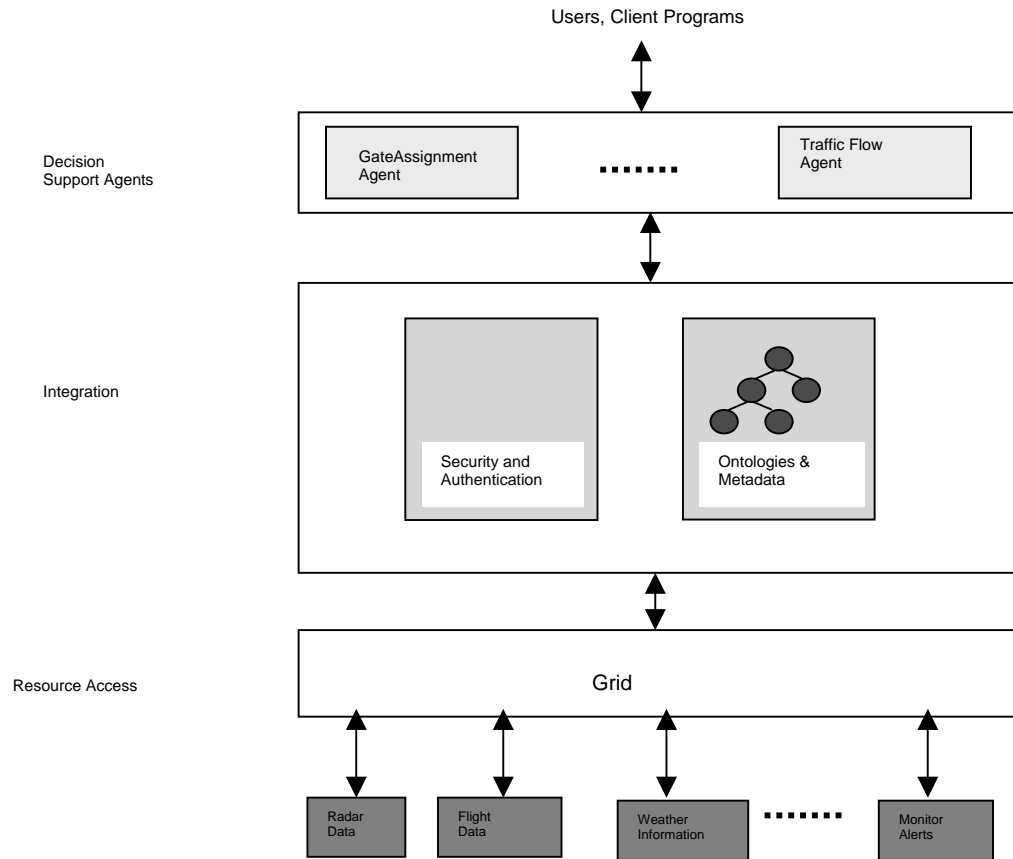

Fig 1. SWIM Proposed Information Integration Architecture

A. Information Synthesis

Within the rapid evolution of the information age we find the view of an intelligent *agent* as a reasoning, or highly knowledgeable specialist is giving way to a new view. This view is that agents are more of a programming framework that may use the aggregation, filtering, cataloging and other information transformation services currently emerging. A couple of examples are Google's search services and IBM's Webfountain project (Cass 2004) that will offer deeper analysis of unstructured data, such as the internet, than is possible now. As a consequence of our research, we divide up the "intelligent agent technology pie" into Model Based Synthesis and Model Based Technology Support.

The reason we have done this is that we feel the "higher ground" to be reached in the next several decades is to have information systems offer up a completely new order of information processing where SWIM's agents play a role in constructing models dynamically and reasoning about them in order to augment the information used in the decision making processes of pilots, air traffic controllers, etc. Much of this *synthesized* information will come from a number of unstructured, semi-structured data sources.

1. Model Based Synthesis

An important problem when dealing with information from multiple sources is that of information *synthesis*. For instance, in a certain situation, we may want reports or descriptions of previous occurrences of that situation (for instance say a bad weather situation which forces aircraft rerouting). We will build on our ongoing synergistic

American Institute of Aeronautics and Astronautics

efforts on intelligent text search and mining in the aviation safety domain, where we developing techniques for intelligent synthesis of semi-structured aviation safety related information residing in multiple sources. We are developing techniques to enable search and synthesis in a significantly more sophisticated manner that is currently feasible by keyword-based search engine technology and information retrieval (IR) systems. Beyond search, we are also investigating techniques to identify and search for behaviour and incident patterns in textual descriptions of aviation related incidents or accidents. Amongst other information, document, textual and semi-structured information will also be important in our application domain, hence we would be addressing the information synthesis issues for such information in this effort.

We must also mention highly relevant in-house technology in this area, the NETMARK data management system (Maluf et al. 2003) which is a system for integrating and managing semi-structured data, that has been developed by the AEN Infolab at NASA Ames Research Center. NETMARK has already been used for several important NASA applications and is a promising technology for addressing some of the important TSWIM objectives as well.

2. Model Based Technology Support
Intelligent agent technology use has been growing significantly in industrial enterprises to the point that this technology is becoming main stream and in many ways the distinction between agents and portions of an operating system, middle layer component or application plug-in is being blurred. As a consequence, agent technology may be viewed as a programming framework that supports intelligent mechanisms based on algorithms and techniques found in artificial intelligence. With the advent of the Internet age the need to deal in semantics, i.e. ontologies, in order to manage the huge explosion of information confronting enterprises the role of agents is more of consumers of semantic information, i.e. metadata than as the main custodians of it, which was the predominate case just a few years ago. This is not to say that significant progress is still under way in making agent technology an order of magnitude more effective. There are several areas that make or break an agent framework's suitability.

The first one is agent habitat. By habitat we mean the services and environment that the agent framework must have in order to operate. Closely related is the agent's framework capable to use feature in its habitat to support its existence and to fulfill is role or behaviors. This boils down to the richness of services included within the framework to deal the kind of computing environment in which the agents are embedded or meant to service.

The second important suitability factor for an agent framework is the richness of its internal representation of knowledge. Knowledge representations include specialized languages that support first and higher order logic constructs and may include some sort of object and its attributes and properties representation. The issue is whether the agent's internal representation of knowledge is sufficient to represent the domain under consideration and how easy it is to use and maintain.

B. Agent and Grid Technologies Survey
Along with the requirements analysis, we also initiated the survey and evaluation of several relevant technologies in the agents, grid and information mediation domains. At this stage we have identified for consideration certain relevant technologies, systems and frameworks; a detailed evaluation and appropriateness determination will be completed later.
The technologies and systems under consideration so far include:

1. SoFAR: An agent framework for distributed information management.(Luc Moreau 2000)
This is a multi-agent framework designed for distributed information management tasks. Has been developed by a research group at the University of Southampton and is available free for research use.

2. Fred Agent Platform (http://www.netdynamics-tech.com)
This is a commercial technology developed by Net Dynamics Internet Technologies GmbH and Co KG. It is an agent platform that enables automated information processing by agents, there is also strong support for semantic web technology and ontologies in this system.

3. Agent Framework for Grid Computing (Majid Ali Khan 2003)
An agent based system for grid computing developed by the University of Central Florida.

4. KAON (http://kaon.semanticweb.org)
An open-source ontology management infrastructure targeted for business applications, developed by the University of Karlsruhe.

5. KaoS (Bradshaw 1996)

An distributed agent development and communication system developed by Boeing and the University of West Florida. Has been used in previous information integration efforts at NASA Ames.

C. Ontologies and Semantic Integration

In the first phase of this task we have the started evaluation of ontology and semantic integration issues to be dealt with as part of the development of AIMS agents and to seek out tools that may aid us in the development of an AIMS ontology base.

The capability of information systems to process information based on semantics, i.e. meaning is a significant challenge; one that is considerably more demanding than mere key, index or textual pattern comparison. The reason for this is that semantic meaning depends on context. To narrow the semantic interpretability space the use of ontologies has come into vogue as a means to more precisely limit the space of possible interpretations, i.e. context; the goal for any given semantic unit within a vocabulary, i.e. ontology, is to have its meaning be resolvable with no ambiguity. While this may be accomplished within a narrow domain with the use of formal methods there remains the issue of information synthesis across multiple ontologies. There is much work under way on this issue. For example see (Maedche et al. 2003) for a tool to manage distributed (distinct) ontologies collectively. This problem does not negate an enterprise's need to invest in the development of its own ontologies and to invest into integration tools for advancing its decision support systems but to be prepared to incorporate the use of semi-structured and unstructured data as part of its ontology portfolio. Emerging technologies such as NASA's NETMARK and IBM's Webfountain, mentioned earlier, address the use of this kind of information. The challenge is to incorporate these as part of an enterprise's collective knowledge base.

D. Security and Access Control Issues

A major practical problem in data sharing and integration across divisions and departments in an enterprise or across multiple enterprises (as will be the case with SWIM) is that of security and access control for the data being shared and integrated. While we can insert data integration technology to facilitate data sharing and integration we also need to ensure that all access control policies for the different pieces of data are indeed being respected. The database research has recently begun to look into this problem (Clifton et al. 2004). Also the industry appears to have made progress too, for instance we have had discussions with Composite Software Inc. a data integration company that claims to have some solutions to the security and data access issues in a data integration environment.

IV.   Operating Realities in the NAS

Meetings with senior personnel at the AATT office at NASA Ames provided us with a valuable introductory insight into some of the non-technical issues and challenges peculiar to the NAS domain, as well as some of the realities of how the airspace is currently managed. As regards information and data, the same data will likely be used by different tools for different tasks and by different roles (personnel). The timescale requirements of the data will likely be very different for different applications. Further, the spatial scale and granularity of data required will be different for different applications. Also the level of detail required can be quite different.

Uncertainty is another big issue in the NAS today. For instance decisions could be taken based on certain assumptions that may not be true (for instance a decision is taken with the assumption that a particular flight will take off at 5pm but it does not). Information may not be completely shared (for instance a pilot has certain information that he does not convey to potentially relevant parties). Further there could be deviations such as an aircraft not being on its filed flight plan (which we are told occurs quite frequently).

The above problems and issues are more political and legacy, rather than of a technical nature. However any solution to information integration and management, such as the AIMS must be developed bearing such factors in mind. To the extent that is feasible, the AIMS team will endeavor to recommend solutions (technical or otherwise) to the above such problems when providing integrated access to information in the NAS domain.

One step the AIMS project will take in the next quarter to manage these requirement issues is to develop a requirements trade space matrix for AIMS and use this matrix to factor out requirements and constraints in the development of agent based DSS tools for SWIM. The scope of this effort, which could be a project in itself, will be subject to limited task resource constraints at hand, but will serve as a model to base subsequent TSWIM projects for deploying agent based information synthesis aids in the NAS and FAA domains.

V. Conclusions

This paper has been an introduction to issues facing program managers and technologists on how to apply semantic-Web technologies to the very broad and diverse domain of aviation. Any effort in providing information integration and fusion in this domain has to work around a number of legacy issues such as information stove-piping and the fact that many databases, data repositories, and real-time operational data streams may have significant incompatibilities that arise when one tries to correlate events and activities across these sources. This problem is sometimes referred to as a "semantic gap" and when the incompatibilities include structural properties in how the data is represented and/or stored then this is sometimes referred to as an "information impedance mismatch".

Our conclusion is that Semantic-Web technologies do have a significant potential for impact on information fusion needs in DDS components of the NAS. No information technology is likely to be a silver bullet and solve all information fusion needs for DDS within the NAS. But with the correct approach in application of these technologies the chance of making a difference for pilots and passengers as well as the flight operations personnel is real. We plan to carry forward this research and develop the AIMS prototype information pool and demonstrate it by linking together existing DDS tools to provide an information fusion capability currently not available

References

Berners-Lee, T., J. Hendler, et al. (2001). "The Semantic Web." Scientific American **184**(5): 34-43.

Bradshaw, J. (1996). KAoS: An Open Agent Architecture Supporting Reuse, Interoperability, and Extensibility. Knowledge Acquisition Workshop, Australia.

Cass, S. (2004). A Fountain of Knowledge. IEEE Spectrum**:** 68-75.

Clifton, C., M. Kantarcioglu, et al. (2004). Privacy-preserving data integration and sharing. DMKD.

Green, S. and J. Jackson (2000). Current and Future Developments in Air Traffic Control. Automation, Control and Complexity: An Integrated View. T. Samad and J. Weyrauch, John Wiley and Sons.

Knoblock, C. A., S. Minton, et al. (2001). "The Ariadne Approach to Web-based Information Integration." International Journal of Cooperative Information Systems (IJCIS) Special Issue on Intelligent Information Agents: Theory and Applications **10**(1/2): 145-169.

Luc Moreau, N. G., David DeRoure, Samhaa El-Beltagy, Wendy Hall, Gareth Hughes, Dan Joyce, Sanghee Kim, Danius Michaelides, Dave Millard, Sigi Reich, Robert Tansley, Mark Weal (2000). SoFAR with DIM Agents: An Agent Framework for Distributed Information Management. The Fifth International Conference and Exhibition on The Practical Application of Intelligent Agents and Multi-Agents, Manchester, UK.

Maedche, A., B. Motik, et al. (2003). "Managing multiple and distributed ontologies on the Semantic Web." VLDB Journal **12**(4): 286-302.

Majid Ali Khan, S. K. V., Kresimir Sivonick and Ladislau L. Boloni (2003). Towards an agent framework for grid computing. Concurrent Information Processing and Computing - Advanced Research Workshop, Sinaia, Romania.

Maluf, D. and P. Tran (2003). NETMARK: A Schema-Less Extension for Relational Databases for Managing Semi-structured Data Dynamically. ISMIS.

NASA (2003). Surface Management System : Operational Concept Description. Moffett Field, NASA Ames Research Center.